

STATISTICS WORKSHEET

* Answers are marked in GREEN

1. Bernoulli random variables take (only) the values 1 and 0.

a) True

b) False

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

a) Central Limit Theorem

b) Central Mean Theorem

c) Centroid Limit Theorem

d) All of the mentioned

3. Which of the following is incorrect with respect to use of Poisson distribution?

a) Modeling event/time data

b) Modeling bounded count data

c) Modeling contingency tables

d) All of the mentioned

4. Point out the correct statement.

a) The exponent of a normally distributed random variables follows what is called the log- normal distribution

b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent

c) The square of a standard normal random variable follows what is called chi-squared distribution

d) All of the mentioned

5. _____ random variables are used to model rates.

- a) Empirical
- b) Binomial
- c) Poisson
- d) All of the mentioned

6. Usually replacing the standard error by its estimated value does change the CLT.

- a) True
- b) False

7. Which of the following testing is concerned with making decisions using data?

- a) Probability
- b) Hypothesis
- c) Causal
- d) None of the mentioned

8. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

- a) 0
- b) 5
- c) 1
- d) 10

9. Which of the following statement is incorrect with respect to outliers?

- a) Outliers can have varying degrees of influence
- b) Outliers can be the result of spurious or real processes
- c) Outliers cannot conform to the regression relationship
- d) None of the mentioned

10. What do you understand by the term Normal Distribution?

Normal distribution represents continuous random variables with approximation or exact values. The normal distribution is the probability function that describes how the values of a variable are distributed.

11. How do you handle missing data? What imputation techniques do you recommend?

Missing data can be dealt with in a variety of ways depending on the need. The most common reaction to it is to ignore it. The techniques that are used to handle missing data are

- Mean Imputation
- Median Imputation
- Mode Imputation
- Hot deck Imputation
- Cold deck Imputation
- Regression Imputation

The methods to handle sometimes can be general or intuitive and can also depend on the domain where we must consult domain expertise to proceed.

12. What is A/B testing?

A/B testing is a method to compare two versions of a single variable, usually by testing a subject's response to variant A against variant B and determining which of the two variants is more effective.

13. Is mean imputation of missing data acceptable practice?

Yes, it is acceptable, imputation of the mean preserves the mean of the observed data. So, if the data are missing completely at random, the estimate of the mean remains unbiased.

14. What is linear regression in statistics?

In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or many explanatory variables also known as dependent and independent variables. The case of one explanatory variable is called simple linear regression, for more than one variable, the process is called multiple linear regression.

15. What are the various branches of statistics?

The two main branches of statistics are descriptive statistics and inferential statistics. Both are employed in scientific analysis of data.